Human Annotation Protocol using a Computer Assisted Method for Validating Electrocardiographic Algorithms to Suppress False Alarms and Reduce Alarm Fatigue

by

Leah Fawcett

THESIS

Submitted in partial satisfaction of the requirements for degree of

MASTER OF SCIENCE

in

Nursing

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Michele Pelter — Michele Pelter

190CFABB136B4BD...

Chair

DocuSigned by:

Lynda Mackin

[illegible]410...

DocuSigned by:

Dr Xiao Hu — Dr Xiao Hu

26AEFBC014E5439...

Committee Members

Acknowledgements:

This Study was done in direct collaboration with:

Dr. Michele Pelter

Dr. Fabio Badilini

Contributions:

Data from this study was retrieved from the data bank from the UCSF Center for Physiologic Research (CPR) and was previously analyzed in the following study:

Drew, B.J., Harris, P., Zègre-Hemsey, J.K., Mammone, T., D. Schindler, R. Salas-Boni, Y. Bai, A. Tinoco, Q. Ding, and X. Hu. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PLoS One*. 2014; 9(10):e110274.

Image: Drew, B.J., Harris, P., Zègre-Hemsey, J.K., Mammone, T., Schindler, D., Salas-Boni, R., Bai, Y., Tinoco, A., Ding, Q., and Hu, X. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PLoS One*. 2014; 9(10):e110274.

Human Annotation Protocol using a Computer Assisted Method for Validating Electrocardiographic Algorithms to Suppress False Alarms and Reduce Alarm Fatigue

Leah I. Fawcett

Abstract

Background: Inundation with alarms in the Intensive Care Unit (ICU) creates alarm fatigue for clinicians and contributes to patient safety events including injury and death because true events are missed. Electrocardiographic (ECG) alarms significantly contribute to this problem because nearly 90% are false positives. These false alarms interrupt care, desensitize clinicians to all alarms, and ultimately reduce responsiveness by clinicians. Research directed towards improving ECG alarm algorithms and alarm specificity is being conducted, however, no current standard for human annotation of true versus false events has been established. **Objectives:** This study was designed to examine annotation methods for analysis of Ventricular Tachycardia (VT) alarms as true versus false and identify challenges for improving these methodologies. The purpose of this study was to: (1) describe a computer-based annotation tool for annotating VT alarms using 24-hour ICU patient files; (2) examine inter-rater reliability between two ICU nurse annotators; and (3) identify specific areas for improving the annotation tool and training preparation of nurse annotators. **Results:** Annotation of 749 VT alarms by two independent ICU nurse reviewers resulted in agreement of 55% (Kappa 0.216). Reviewer 1 identified 333 (44.5%) of the alarms as false, whereas Reviewer 2 identified 605 (80.8%) as false. **Conclusion:** In this pilot study, we identified that more training and education of annotators was needed and that the annotation protocol could be improved by reducing the number of annotation categories, single and random presentation of alarms, rather than an entire 24-hour file for one patient, and validation by an ECG expert with education and training developed from disagreements and incorrect annotations.

Table of Contents

Materials and Methods

List of Figures

List of Tables

Introduction: The problem of Alarm Fatigue

In the hospital setting, alarms from bedside patient monitors are designed to cue clinicians of serious changes in a patient's condition. In the intensive care unit (ICU) specifically, a cacophony of other alarm sounds are also produced by IV pumps, call lights, ventilators, bed alarms, phones, and pagers, which all congest the clinician's ear and distract from patient care. Each alarming device beckons for a response, requiring a trained clinician's eye and clinical expertise to assess and intervene as appropriate. Recent studies have shown that the sheer number of alarms triggered in the patient care setting contributes to a phenomenon called alarm fatigue.[1-3] Alarm fatigue is a term used to describe the sensory overload experienced as a result of incessant inundation from alarms.[4] The psychological effect of alarm fatigue in nurses is desensitization towards alarms.[5] Desensitization has been shown to cause dangerous behaviors such as delayed responsiveness, silencing of alarms without assessing the patient, lowering the alarm volume or silencing altogether, and not hearing alarms (alarm fatigue) potentially leading to missed true events.[2, 6] Perpetuating the problem of alarm fatigue even further, is the fact that the majority of these alarms will be deemed false, [5, 7] which is defined as an alarm that is inaccurate or nonactionable.[8]

Electrocardiographic (ECG) alarms have been shown to be particularly problematic. In the most comprehensive study to date, of 12,671 audible ECG arrhythmia alarms annotated in 461 consecutive ICU patients, 88.8% were determined to be false.[9] Of the arrhythmia types annotated, ventricular tachycardia (VT) were the second most frequent (3,861/12,671; 31%), and 86.8% of the VT alarms were annotated to be false. Because true VT often necessitates urgent patient intervention, VT is configured as a "crisis" level alarm in most bedside monitors. Crisis alarms are designed to “latch” or beep continuously at a loud, distinct high pitch until the nurse physically silences the alarm. Latching alarms create an even greater problem with regards to alarm fatigue because they require the clinician, in most cases a nurse, to silence them. This results in an interruption in patient care because the nurse must not only silence the alarm, but also quickly assess whether the alarm is true or false.

Alarm fatigue creates a significant patient safety problem.[11] One study examined response times by nurses and showed that as the number of alarms increased during a nurses shift, so did the response time.[11] Behaviors in response to alarm fatigue by nurses and others responsible for overseeing alarms include: 1) silencing alarms without assessing the patient 2) lowering the alarm volume, 3) permanently disabling alarms, and or 4) delayed responsiveness to alarms when presumed false.[8] These actions place patients at risk for serious adverse events, including death, because true alarms are missed.

Alarm fatigue has been identified by regulatory bodies, safety, and nursing organizations as a serious patient safety issue. In 2013, the Joint Commission established national patient safety goal #6 “Reduce Harm Associated with Clinical Alarms.[12]” This came about after a Sentinel Event Alert that revealed 98 patient incidents related to alarm fatigue, of which 93, resulted in a patient death or permanent loss of function.[13] The Emergency Care Research Institute has placed alarm fatigue at or near the top of its top ten list of patient safety hazards since 2010.[6] The American Nurses Association and the American Association of Critical-Care Nurses have issued practice alerts regarding alarm fatigue, emphasizing the need for evidence-based approaches to solve this complex problem.[3, 14, 15]

In a review of literature surrounding interventions to reduce alarm fatigue, recommendations include: adjusting default monitor presets (i.e. vital sign parameters, Sp02 parameters, respiratory rate parameters, etc.), daily skin electrode changes, frequent alarm customization for individual patients, and alarm management education.[4, 12] However such interventions are aimed at addressing human factors that contribute to alarms (i.e., motion artifact, noise, or exceeding vital sign parameters), but fall short of addressing the primary causal factor of false alarm – that is poorly designed arrhythmia algorithms.[4, 8, 9]

ECG monitoring systems use alarm suppression algorithms that filter data input prior to triggering an alarm. Such algorithms vary in effectiveness of suppressing false alarms. A systematic review by Winters et al. reported a variation from 20% to 100% in effectiveness of false alarm suppression in ECG algorithms, with little to no effect on suppression of true alarms.[4] However, one study showed their algorithm also suppressed true VT (17.3%) and asystole (2.33%). These studies suggest that alarm suppression algorithms may be a promising approach to reduce false alarms if the threat to patient safety is minimized.[16-20] Recent efforts to improve ECG algorithms designed for increased alarm specificity without decreased sensitivity have begun. However, in order to ensure true and false alarms are identified accurately, algorithms must be carefully validated by clinical experts prior to seeking approval from regulators of monitoring devices (i.e., US Federal Drug Administration). However, there currently exists no established standardized annotation protocol for evaluation of newly designed ECG algorithms, leaving the opportunity for great variations. Challenges to large-scale annotation efforts include the potential for variability between annotators, the labor-intensive nature of manual annotation, ensuring expertise of annotators, obtaining data of adequate sample size, and access to computerized tools for annotating large amounts of data. Addressing these potential barriers could have a major impact in achieving high quality annotation aimed at addressing alarm fatigue.

For the past several years, our team has developed web-based tools and approaches that have used human annotation to test algorithms for code blue events, arrhythmias, and myocardial ischemia.[21-27] From these studies, we have learned that it is extremely difficult to secure physician annotators due to the demand on their time in clinical practice. Hence, we believe highly trained and expert nurses who work closely with ECG data could perform the initial annotation of ECG alarms. Physician annotators could then be utilized to validate these efforts using a smaller sub-set of data (i.e., random sampling of data, and disagreements). As such, this study was designed to address the following aims: (1) test a computer-based annotation tool for annotating true/false VT alarms from 24-hour ICU patient files; (2) examine inter-rater reliability between two ICU nurse annotators, and (3) identify specific areas for improving the annotation tool and training preparation for nurse annotators.

Materials and Methods

Sample and Setting

This was a secondary analysis using ECG data collected from the UCSF Alarm Study, which has been described in detail previously.[9] Briefly, the UCSF Alarm Study included all consecutive patients (n = 461) admitted to one of three ICU types (i.e., cardiac, medical/surgical, neurological). The UCSF Committee on Human Research approved the study with waiver of patient consent because all ICU patients have physiologic monitoring as part of their routine care and acquisition and storage of these data, which was examined on a secure server retrospectively in our research lab, did not influence clinical care. For the present study, we included any patient with a VT alarm (true or false) during continuous ECG monitoring during a one-month period in March 2013. In addition to continuous ECG waveform data, all physiologic waveform data were obtained if the patient had them (i.e., SpO2, arterial blood pressure), using our sophisticated research infrastructure (Figure 1).[9] The sample available for the present pilot study were a subgroup of 63 patients that produced 749 VT alarms that were annotated by two ICU nurses.

Instruments and Procedure

A computer-based annotation tool was designed for this pilot project that presented a 24-hour time period of de-identified ECG data for each patient. Because it is not uncommon for critically ill patients to be in the ICU for several days, we chose to parse each patient file into 24-hour time periods. The 24-hour time files made it easier for annotators, who may have to annotate hundreds of VT events during one 24-hour period, but also for tracking events for future analyses. We chose to start a 24-hour time period at midnight and ended each time period at 11:59:59. If a patient was admitted at 3:00 AM for example, their 24-hour time period started at 3:00 AM and ended at 11:59:59. If this same patient was in the ICU for another 24-hour period, their second file started at midnight and ended at 23:59:59 the following day, or earlier of they were discharged prior to this time. Files for all patients with multi-day admissions were created using this same format.

Annotation of the alarms for this pilot study was performed using a software program developed by our co-author (FB; CER-S, AMPS-LLC, New York) which was installed onto two workstations in the UCSF Center for Physiologic Research (CPR). Each workstation had a 24-inch monitor for viewing the VT event. This large screen size allowed for excellent vision of the ECG and physiologic waveforms. The CER-S annotation tool uses all of the ECG and physiologic waveform data acquired from the bedside monitor. This is typically five ECG leads (I, II, III, aVR, aVL, aVF and one V lead; our hospital uses V1), waveform data (i.e., arterial blood pressure, intracranial pressure, etc.), transthoracic impedance for respiratory rate, and Sp02 photoplethysmogram (PPG). For this study, we also included four derived chest leads to aid with annotation. All of the available signals are displayed in a cascade together with all the alarms triggered by the monitor as depicted in Figure 2.

For this project, only VT alarms from our database were annotated. After secure password-controlled access, the annotator could visualize the by-patient list of VT alarms. Through a simple graphic interface, several choices for the VT event using a drop-down menu were presented to the annotator. The dropdown list included the following options: False, True, Single Lead, Cardiopulmonary Resuscitation (CPR), Electronic Pacing, Atrial Tachycardia w/ Wide QRS, Polymorphic VT, Sustained VT, Irregular VT, Torsades de pointes, and Unsure. These annotation labels provided a sub-categorization for VT alarms that were categorized as “True” or “False” As shown in Table 1.

Annotation Protocol

Two nurses were designated to annotate VT alarms. The annotators are both experienced ICU nurses with several years of clinical experience in the ICU setting with educational training for identifying lethal arrhythmias including VT. One nurse has earned a Master's degree and the other is currently in graduate school. Prior to performing the annotations, each reviewer was trained individually to use the CER-S software by an expert (MMP). This session covered not only the basics of how to use the CER-S tool, but identified challenging annotations. These cases were then used as examples at a team meeting to help guide and develop decisions about true/false categories prior to the formal annotation effort.

This one-on-one training was followed by a meeting of the entire team (i.e., engineers of the CER-S tool, nurse annotators, and ECG analytics expert) to discuss definitions for true versus false VT (Table 1). This meeting re-enforced how to use the CER-S tool as well as how to annotate more complex VT alarms. For example, if only a single lead was present, the annotator was to select "Single Lead," or if CPR was present the annotator should select "CPR." At the conclusion of the meeting, it was approximated that each annotation would take on average 60 seconds to complete.

Process

Annotators entered the UCSF CPR Lab, which required I.D. badge access to ensure security and confidentiality. Each annotator was provided with an individualized passcode to log onto one of two computer workstations that contained the CER-S application tool. The login meant the annotator viewed only records assigned to them, and only ICU patient files with one or more VT alarms were assigned. Once a file had been annotated, the annotator closed the patient file and was prompted to “save yes/no” the file. If yes was selected, the patient file disappeared from the annotator’s database. After login, annotators opened the CER-S program from the desktop and were immediately presented with a list of all patients with ≥ 1 VT alarm needing annotation. Each admission was coded with a random identification number, which de-identified the file from any private patient information (i.e., demographics, dates, etc.). Each separate admission file had the total number of VT alarms triggered during the 24-hour (or less) monitoring period.

Diagnosis of True versus False Alarms

The criteria for defining VT (Table 1) were developed from standard definitions for VT as establish by the Heart Rhythm Society (HRS), American College of Cardiology (ACC), American Heart Association (AHA) and those used in previous alarm annotation studies.[9] VT definition criteria were posted at the annotation workstations for reference.

Statistical Analysis

The CER-S tool is designed so that at the end of the reviewing process a script-based program is executed by the project supervisor to automatically export the results for each reviewer into a spreadsheet file format. We used Excel for this initial step, which contained all of the annotated alarms and the manual selections made by each reviewer. The Excel spreadsheet was then imported into a statistical program (SPSS 25.0;IBM Corporation, 2017) for analysis of inter-rater reliability. After a three-week annotation period, the data for the two annotators was exported into SPSS for analysis of inter-rater reliability assessment.

Descriptive Statistics (frequencies and proportions) were used to report the frequency of alarm annotations by each annotator. A Kappa score for agreement between the two annotators was also calculated. Frequency distributions were also examined between the two reviewers to examine the distribution of selections for each annotation category. A cross tabulation of agreement was used to display a percentile of agreement between Reviewer 1 and Reviewer 2’s annotations.

Results

A total of 749 VT alarms were annotated by the two reviewers (Table 2). Of these, Reviewer #1 annotated 333 (44.5%) of alarms as false, whereas Reviewer #2 annotated 605 (80.8%) of the alarms as false. Figures 3, 4, 5, and 6 are examples of VT alarms where there was disagreement between the two reviewers. As shown in Table 3, there was agreement in 410 of the 749 VT annotations (55%). Also seen in Table 3, a total of 34 of the 338 disagreements were annotated as False by Reviewer 1 and True by Reviewer 2 (4.5%), whereas 304 of the 338 disagreements were annotated true by Reviewer 1 and false by Reviewer 2 (40%). The calculated Kappa statistic for agreement between annotations by the two reviewers was 0.216.

Discussion

This study established that the developed CERS tool could display VT alarms for individual patients in a 24-hour file format without difficulty. Annotation selections for each reviewer were properly stored and were in a format that allowed for transfer into Excel and then a statistical package for analysis. We estimated that annotations would require less time to complete (approximately 60 seconds/alarm), but were much longer in length presumably due to too many annotation categories to choose from, rather than a simple true or false choice. Due to time constraints, education and training were short and may have also contributed to low inter-rater reliability.

The results of this study show there was low agreement between annotators examining VT alarms. We have identified several challenges and potential confounders that can occur between reviewers performing annotation. Table 4 lists the identified challenges and confounders we identified following an examination of disagreements in this study.

Conclusion

In studies examining the cause of ECG alarms and an evaluation of new algorithms to reduce false alarm incidence, annotations must be done by clinicians to ensure accuracy. Superior future study designs should involve multi-rater blinded review of ECG alarms for annotation of true/false utilizing a computerized application with capacity for assessing a patient's baseline ECG rhythm and corresponding waveform tracings (ABP, sp02, respiratory rate, etc.) for time periods before and after the triggered alarm. Since manual annotation provides the potential for human factors such as bias and subjectivity, this presents a serious confounder to the validity of such annotation efforts. These confounders can be mitigated by: ensuring sufficient time for education and pilot testing; making adjustments following a pilot-study; sampling data so as to represent a variety of patients; randomization of alarm presentation; and reducing categories for alarm annotation selection. Future research is indicated to validate the effectiveness of this protocol and to quantify the influence on agreement between reviewers. Such evidence could be utilized to validate methods of annotation and analysis through the evolution of technology and algorithms for alarm fatigue reduction.

References

1. Kowalczyk, L. Alarm fatigue a factor in second death. *Boston Globe*. September 21, 2011. www.bostonglobe.com/2011/09/20/umass/qSOhm8dYmmaq4uTHZb7FNM/story.html. Accessed April 24, 2019.
2. Sendelbach, S. and Funk, M. Alarm fatigue: a patient safety concern. AACN Adv Crit Care. 2013;24(4): p. 378–386.
3. Chopra, V. and McMahon, L.F. Redesigning Hospital Alarms for Patient Safety: Alarmed and Potentially Dangerous. *JAMA*. 2014;311(12):1199–1200. doi:10.1001/jama.2014.710
4. Winters, B.D., Cvach, M. M., Bonafide, C.P., Hu, X., Konkani, A., O'Connor, M. F., Rothschild, J.M., Selby, N. M., Pelter, M. M., McLean, B., Kane-Gill, S. L., A Society for Critical Care Medicine, and Alert Fatigue Task, Technologic Distractions (Part 2): A Summary of Approaches to Manage Clinical Alarms With Intent to Reduce Alarm Fatigue. *Critical Care Medicine*; 2018; 46(1): p. 130-137.
5. Lawless, S.T. Crying wolf: false alarms in a pediatric intensive care unit. *Critical Care Medicine*. 1994 Jun;22(6): p. 981–1005.
6. ECRI Institute. Top 10 health technology hazards. *Health Devices*. Retrieved from https://www.ecri.org/Pages/2018-Hazards.aspx
7. Tsien, C.L. and Fackler, J.C. Poor prognosis for existing monitors in the intensive care unit. *Critical Care Medicine*. 1997 Apr;25(4): p. 614–619.

8. Harris, P.R., Zègre-Hemsey, J.K., Schindler, D., Bai, Y., Pelter, M.M., Hu, X. Patient characteristics associated with false arrhythmia alarms in intensive care. *Therapeutics and Clinical Risk Management*. 2017;13: p. 499-513.
9. Drew, B.J., Harris, P., Zègre-Hemsey, J.K., Mammone, T., Schindler, D., Salas-Boni, R., Bai, Y., Tinoco, A., Ding, Q., and Hu, X. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PLoS One*. 2014; 9(10):e110274.
10. Pelter, M., Fidler, R., and Hu, X. Research: Association of Low-Amplitude QRSs with False-Positive Asystole Alarms. *Biomed Instrum Technol*, 2016; 50(5): p. 329-335.
11. Bonafide, C.P., Lin, R., Zander, M., Graham, C.S., Pain, C.W., Rock, W., Rich, A., Roberts, K.E., Fortino, M., Nadkhami, V.M., Localio, A.R., and Keren, R. Association between exposure to nonactionable physiologic monitor alarms and response time in a children's hospital. *Journal of Hospital Medicine*. 2015;10(6): p. 345-351.
12. The Joint Commission. (2014). The Joint Commission announces 2014 National Patient Safety Goal. *Joint Commission Perspectives*. 33(7): p. 1-4. https://www.jointcommission.org/assets/1/18/JCP0713_Announce_New_NSPG.pdf
13. The Joint Commission. Patient Safety Alert: Medical device alarm safety in hospital, in Sentinel Event Alert. 2013. (50) p. 1-3.
14. Do you hear what I hear? Combating alarm fatigue. *American Nurses Association*. 2014;9(11).

Retrieved from https://www.americannursetoday.com/hear-hear-combating-alarm-fatigue/

15. AACN Practice Alert. Managing Alarms in Acute Care across the Life Span: Electrocradiography and Pulse Oximetry. *Critical Care Nurse*. 2018;38(2): p. 16-20.
16. Salas-Boni, R., Bai, Y., Harris, P.R., Drew, B.J., and Hu, X. False ventricular tachycardia alarm suppression in the ICU based on the discrete wavelet transform in the ECG signal. *Journal of Electrocardiology* 2014; 47: p. 775–780.
17. Baumgartner, B., Rodel, K., and Knoll, A.: A data mining approach to reduce the false alarm rate of patient monitors. *IEEE Engineering in Medicine and Biology Society* 2012; p. 5935–5938.
18. Desai, K., Lexa, M., and Matthews, M. and Genc, S.: Hemodynamic impact based prioritization of ventricular tachycardia alarms. *IEEE Engineering in Medicine and Biology Society* 2014; p. 3456–3459.
19. Aboukhalil, A., Nielsen, L., Saeed, M., Mark, R.G., and Clifford, G.D.: Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. *Journal of Biomedical Informatics* 2008; 41: p. 442–451.
20. Li, Q. and Clifford, G.D.: Signal quality and data fusion for false alarm reduction in the intensive care unit. *Journal of Electrocardiology* 2012; 45: p. 596–603.
21. Bai, Y., Do, D., Ding, Q., Palacios, J.A., Shahriari, Y., Pelter, M.M., Boye, N., Fidler, R., and Hu, X. Is the Sequence of SuperAlarm Triggers More Predictive Than Sequence of the Currently Utilized Patient Monitor Alarms? *IEEE Transactions on Biomedical Engineering*. 2017;64(5): p. 1023-1032.

22. Bai, Y., Do, D.H., Harris, P.R., Schindler, D., Boyle, N.G., Drew, B.J., and Hu, X., Integrating monitor alarms with laboratory test results to enhance patient deterioration prediction. *Journal of Biomedical Informatics* 2015;53: p. 81-92.
23. Salas-Boni, R., Bai, Y., Harris, P.R., Drew, B.J, and Hu, X., False ventricular tachycardia alarm suppression in the ICU based on the discrete wavelet transform in the ECG signal. *Journal of Electrocardiology*, 2014;47(6): p. 775-780.
24. Shahriari, Y., R. Fidler, M. Pelter, Y. Bai, A. Villaroman, and X. Hu, Electrocardiogram Signal Quality Assessment Based on Structural Image Similarity Metric. *IEEE Transactions on Biomedical Engineering*, 2017; 65(4): p. 745-753.
25. Gadhoumi, K., Do, D., Badilini, F., Pelter, M. M., and Hu, X. Wavelet leader multifractal analysis of heart rate variability in atrial fibrillation. *Journal of Electrocardiology*, 2018. 51(6S): p. S83-S87.
26. Pelter, M.M., Xu, Y., Fidler, R., Mortara, D.W., and Xiao, H. Evaluation of ECG algorithms designed to improve detect of transient myocardial ischemia to minimize false alarms in patients with suspected acute coronary syndrome. *Journal of Electrocardiology*, 2018. 51(2): p. 288-295.
27. Xiao, R., Xu, Y., Pelter, M.M, Fidler, R., Badilini, F., Mortara, D.W., and Hu, X. Monitoring significant ST changes through deep learning. *Journal of Electrocardiology*, 2018. 51(6S): p. S78-S8.

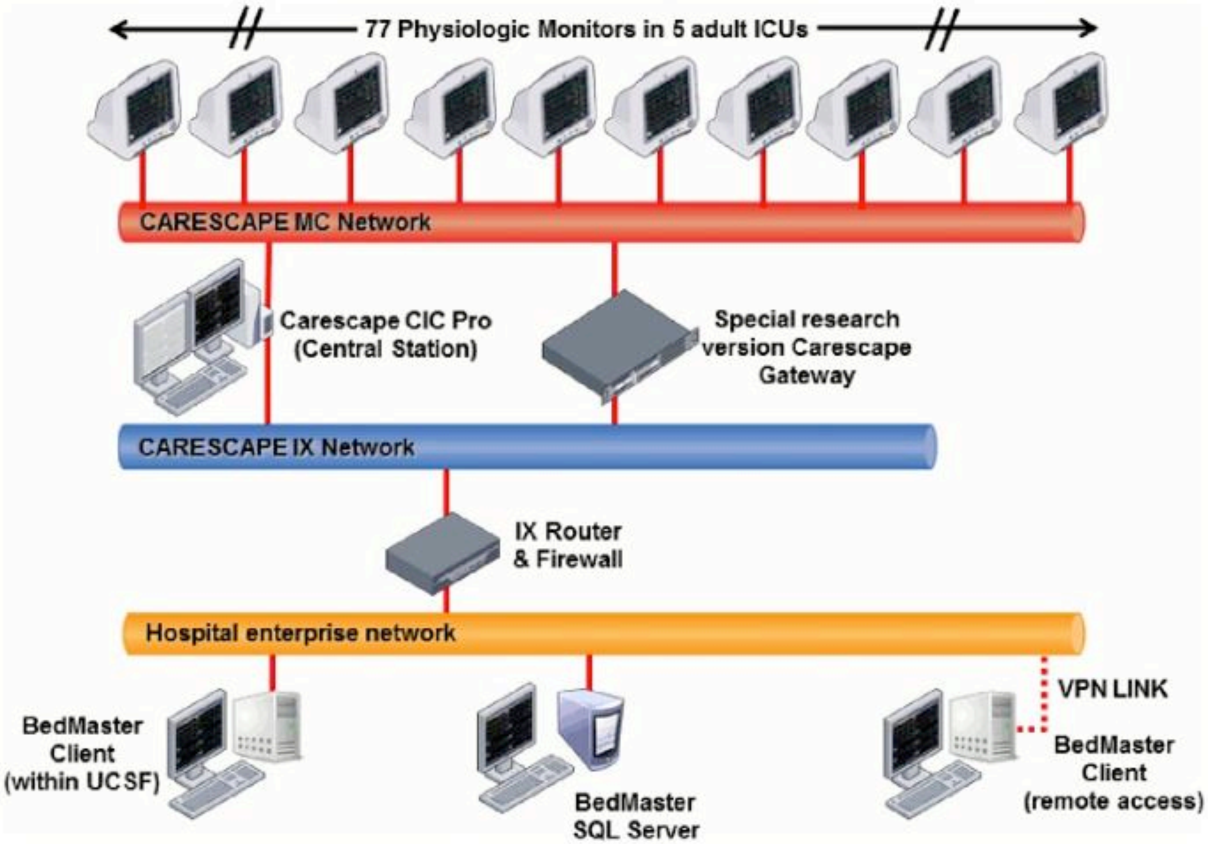

Figure 1. Monitoring infrastructure utilized to store alarm data from bedside physiologic monitors in five adult intensive care units. **Reference for image**: Drew, B.J., Harris, P., Zègre-Hemsey, J.K., Mammone, T., Schindler, D., Salas-Boni, R., Bai, Y., Tinoco, A., Ding, Q., and Hu, X. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PLoS One*. 2014; 9(10):e110274.

```
Lethal Ventricular Arrhythmia: 131
  V Tach: 131
    VTach [3s ] - 07-Mar-2013
    VTach [3s ] - 07-Mar-2013 08:47:48.400
    VTach [3s ] - 07-Mar-2013 08:56:36.400
    VTach [3s ] - 07-Mar-2013 09:00:52.400
    VTach [3s ] - 07-Mar-2013 09:13:06.400
    VTach [3s ] - 07-Mar-2013 09:14:46.400
    VTach [3s ] - 07-Mar-2013 09:17:32.400
    VTach [3s ] - 07-Mar-2013 09:17:40.400
    VTach [3s ] - 07-Mar-2013 09:19:18.400
    VTach [3s ] - 07-Mar-2013 09:37:27.400
    VTach [3s ] - 07-Mar-2013 09:37:49.400
    VTach [3s ] - 07-Mar-2013 09:37:57.400
    VTach [3s ] - 07-Mar-2013 09:38:15.400
    VTach [3s ] - 07-Mar-2013 09:40:11.400
    VTach [3s ] - 07-Mar-2013 09:42:53.400
    VTach [3s ] - 07-Mar-2013 10:05:21.400
    VTach [3s ] - 07-Mar-2013 10:22:11.400
    VTach [3s ] - 07-Mar-2013 10:23:27.400
    VTach [3s ] - 07-Mar-2013 13:10:21.400
    VTach [3s ] - 07-Mar-2013 16:26:24.400
    VTac [3s ] - 07-Mar-2013 19:58:06.000
    VTach [3s ] - 07-Mar-2013 20:05:19.200
    VTac [3s ] - 07-Mar-2013 20:12:39.000
    VTach [3s ] - 07-Mar-2013 20:14:07.200
    VTac [3s ] - 07-Mar-2013 20:15:14.000
    VTac [3s ] - 07-Mar-2013 20:16:47.000
    VTach [3s ] - 07-Mar-2013 20:24:31.700
    VTach [3s ] - 07-Mar-2013 20:26:12.700
    VTach [3s ] - 07-Mar-2013 20:27:14.400
    VTac [3s ] - 07-Mar-2013 20:27:31.000
    VTac [3s ] - 07-Mar-2013 20:27:42.000
    VTach [3s ] - 07-Mar-2013 20:27:57.200
    VTach [3s ] - 07-Mar-2013 20:28:09.200
    VTach [3s ] - 07-Mar-2013 20:28:12.200
    VTach [3s ] - 07-Mar-2013 20:28:19.200
    VTach [3s ] - 07-Mar-2013 20:28:24.200
    VTac [3s ] - 07-Mar-2013 20:28:43.000
    VTach [3s ] - 07-Mar-2013 20:40:18.400
    VTac [3s ] - 07-Mar-2013 20:43:58.000
    VTac [3s ] - 07-Mar-2013 20:44:39.000
    VTach [3s ] - 07-Mar-2013 20:47:26.200
    VTac [3s ] - 07-Mar-2013 20:48:47.000
    VTach [3s ] - 07-Mar-2013 20:50:48.400
    VTach [3s ] - 07-Mar-2013 20:54:16.400
    VTach [3s ] - 07-Mar-2013 20:55:52
    VTach [3s ] - 07-                .400
    VTach [3s ] - 07-             7:10.400
    VTach [3s ] - 07-             8:18.400
    VTach [3s ] - 07-             9:14.400
    VTach [3s ] - 07-Mar-2013 21:10:02.400
    VTach [3s ] - 07-Mar-2013 21:11:12.400
    VTac [3s ] - 07-Mar-2013 21:12:05.000
    VTach [3s ] - 07-Mar-2013 21:12:42.400
```

Figure 2. CER-S Tool. Pictured above is a screenshot view of the CER-S tool when a 24-hour (or less) patient file is opened. This example contains a total of 131 ventricular tachycardia (VT) alarms generated by a single patient within a 24-hr time period. The reviewer is presented with the individual VT alarms triggered (arrow 1) and to the right a screen that displays the first alarm highlighted in pink to the right in a 10-second window of time on an ECG tracing. Each alarm can be selected for viewing and then annotation. The pop-up window for annotation (arrow 2) automatically appears when an alarm is opened. An alarm disappears from the list (arrow 1) once the annotation is performed and is stored in a delineated file for later analysis. The CER-S annotation tool also displays continuous data before and after the VT alarm in a 60-second window (Arrow 3). A vertical scrollbar (arrow 4) allows the annotator the ability to scroll to other corresponding leads and waveform(s) if available (Sp02, arterial blood pressure, central venous pressure [CVP], etc). A horizontal scrollbar (arrow 5) provides the capacity to scroll back and forward in time. When clicking on the main screen where the VT alarm is displayed, a red shadowed line identifies the ECG time period responsible for triggering an alarm (+/- 6 seconds) and is shown with arrow 6. Finally, arrow 7 illustrates where several tools are located that allow the annotator to individualize the screen view. For example, one can change the ECG lead order, measure time and or amplitude with calipers, etc.

Figure 3. Example: Reviewer 1 annotated True and Reviewer 2 annotated False. The ECG tracing pictured above is an example of disagreement between reviewers in this study. Although lead I shows little to no change in QRS morphology, in leads II, III, V, and the remaining chest leads a wider a more rapid QRS complex occurs for 6 beats. The change in polarity and morphology for six beats is marked by a graphic in V lead. Here is a classic example of what one reviewer may have not been considered fast enough or long enough (marked first beat as native versus fusion), whilst another reviewer considered this to be 6 beats of VT. It should be noted that this change in morphology is so slight that it might be missed at a cursory glance of the ECG strip. As displayed by the graphic above, attention should be given to the snapshot view as denoted by the smaller graphic below the ECG tracings in which changes in speed and morphology can be easily detected.

Figure 4. Example: Reviewer 1 annotated True and Reviewer 2 annotated False. The ECG tracing above is another example of reviewer disagreement. In this case, 13 total alarms were triggered from this patient, of which reviewers were in agreement in all but the alarm pictured above. This example displays how both artifact and pacemaker made the ECG tracing difficult to assess. Although pacer spikes can be noted above, variations in QRS speed, direction, and morphology are likely to contribute to confusion for reviewers. The 10-second view (upper image with pink graphed background) displays the data that triggered the alarm as denoted by the red graphic as well as subsequent ECG data. The 60-second snapshot view (pictured horizontally below with a white background) displays the alarm triggering data with a smaller red graphic.

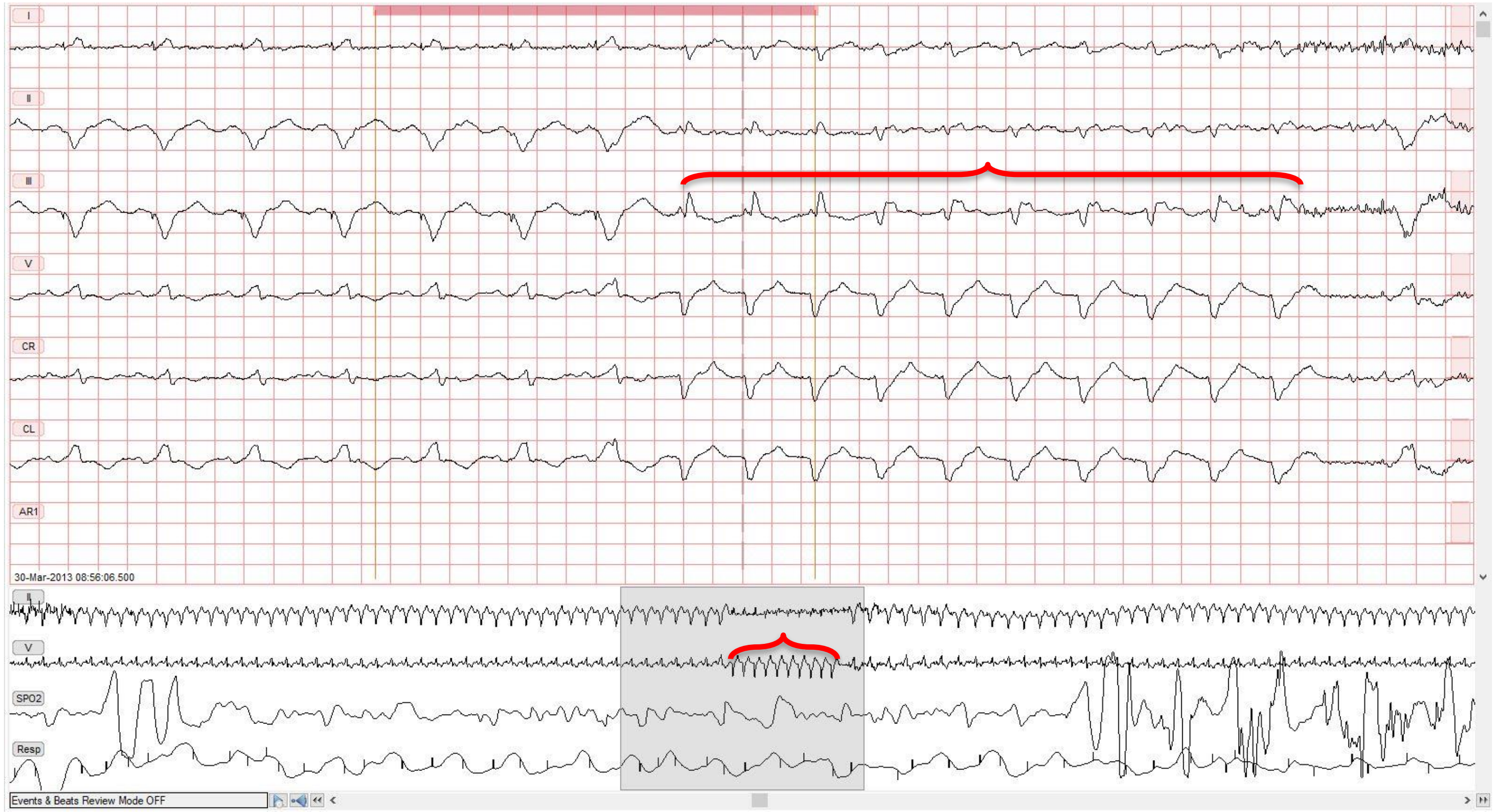

Figure 5. Example: Reviewer 1 annotated True and Reviewer 2 annotated False. The ECG tracing above is another example of reviewer disagreement in this study. This alarm was triggered by a patient with 29 total alarms, of which only one resulted in disagreement between reviewers. Pacer spikes seen in only in lead III, are denoted with an arrow, showing that the patient has an underlying paced rhythm. As highlighted with the graphic in the V lead in the 60-second window in the lower screen, the QRS complex changes in morphology from the baseline, in width, directionality, and rate for a series of 9 beats. It should also be noted that there is no longer a visible pacer spike in lead III during the 9 sequential beats of concern for VT. This instance of disagreement is one of many that are difficult to explain. It should be noted that this alarm was the last of 29 total alarms for this patient; all other alarms were annotated as False by both reviewers.

Figure 6. Example: Reviewer 1 annotated True and Reviewer 2 annotated False. The ECG tracing above is an example of disagreement between reviewers from this study. This patient triggered a minimum of 131 alarms from which many disagreements arose between reviewers. Multiple contributory factors may have caused the disagreements such as: 1) an abnormal baseline rhythm, 2) frequent ectopic beats (as depicted by arrows in Lead II) may have confused reviewers in alarms such as this as to which beats represented the patient's native ECG rhythm, and 3) unusual morphology of the 6 beats in question (depicted with graphics above in Lead III). This depicts how a single difficult-to-analyze patient rhythm can skew agreement results between reviewers if a large amount of total alarms reviewed can be attributed to such a patient. Reviewer 1 clearly considered the complexes denoted by the red arrows to be the native rhythm and therefore the run of 6 beats to be a VT, due to change in morphology and the rate of >100bpm on average, meeting the qualifiers for VT as per definition. Reviewer 2 may have considered the amplitude to be low and the QRS complex too narrow to consider this VT.

Table 1. VT Annotation Definitions.

False VT	• noise induced misinterpretation • average rate below 100 beats per minute • fewer than six ventricular beats
True VT	• At least six consecutive wide QRS complexes • A change of QRS morphology from the preceding rhythm • A QRS which appears to be a fusion of the ventricular beats and the underlying rhythm count toward the 6-beat threshold • At least 2 preceding “normal” beats if a separate VT occurs within preceding 10 seconds • Generally, the rate of the VT should be higher than that of the preceding rhythm. “VT” with a rate lower than the preceding rhythm may reflect electronic ventricular pacing with low visibility of the pacing spikes • The VT episode can occur anywhere within +/- 10 seconds of the recorded alarm time (center of display window)
Irregular VT	• VT with significant variation of RR intervals. The range of RR intervals (longest v. shortest) should exceed 200 milliseconds
Polymorphic VT	• VT with significant QRS morphology variation. Fusion beats, however, should be discounted when considering these variations. • The irregular VT annotation takes precedence over the polymorphic VT when both features are present.
Sustained VT	• VT lasting 12 beats or more, with no more than 1 normal beat sandwiched among the 12 beats • The alarm onset is at least 6 beats after the actual VT onset.
AT w/ wide QRS	• A wide QRS rhythm meeting VT criteria, but with strong evidence that the true rhythm is an atrial tachycardia. • Atrial flutter may also be a source of this rhythm annotation.
Torsades de pointes	• VT with a smooth variation of the QRS

Table 2. Ventricular tachycardia (VT) events as annotated by Reviewer #1 and Reviewer #2.

Annotation	**Reviewer #1**	**Reviewer #2**
False	333 (44.5%)	605 (80.8%)
True	364 (48.5%)	91 (12.1%)
Unsure	2 (0.3%)	4 (0.5%)
Single Lead/CPR	50 (6.7%)	49 (6.5%)
Total	749	749

Table 3. Agreement of annotations between Reviewer #1 and Reviewer #2.

Annotation	**Frequency**	**Percent**
Reviewer 1 False Reviewer 2 True	34	4.5
Agreement	412	55
Reviewer 1 True Reviewer 2 False	306	40
Reviewer 2 True Reviewer 1 Single Lead	1	0.1
Total	749	100

Table 4. ECG Annotation Research Challenges and Confounders with Recommendations

Potential Challenges and Confounders	**Recommendation**
Education and Training: The short timeframe imposed by a required deadline meant education and training may not have been sufficiently long enough.	**Allow for Adequate Time:** Several weeks of education and training are required prior to full-scale annotation. Pilot test annotation and make adjustments to annotation criteria, and the annotation tool based on findings.
Time Ordered/Patient Ordered Data Sample: Annotator bias created because multiple alarms for one patient are in one large file. Example: Annotator may decide the alarm is "false/true" early in a file and then stay with this choice for all remaining alarms in a single patient file.	**Randomized Data Sample:** 1. Annotate a single alarm only. 2. Randomize order of presentation of alarms in order to blind annotators as to which patient produced the alarm as well as to the sequence order of alarms produced in time. 3. Remove Timestamp from application data to blind annotator from time sequence of annotations.
Excessive Categorization: Having too many category codes meant more time and thought was required for annotation. For example, our protocol included the following seven choices: 1. Monomorphic VT 2. Polymorphic VT 3. Noise induced 4. Pacemaker induced 5. CPR 6. Single lead 7. Unsure	**Two Initial Categories for Annotation**: Provide Minimal Options for initial annotation to simplify basic agreement of true/false between annotators. If desired, provide an additional drop down categorizations after initial annotation. For example, utilize a second level of annotation after the initial True/False annotation of: 1. True VT Alarm a. Monomorphic VT b. Polymorphic VT 2. False VT Alarm a. Noise induced b. CPR induced
Many alarms by Few Patients Sample: Data Set includes data from only a small number of patients producing a large number of alarms. Consistent annotation of a single patient's alarm pattern (responsible for a high volume of triggered alarms) in a certain way may skew data. Figure 6 is an example.	**Diversified Sample Selection**: Select pool of alarms that results from a greater number of patients in order to reduce potential for a few patients with multiple alarms that will skew data analysis.

Chest Leads: Inclusion of “Chest Leads” in the first screen view, rather than the ECG leads use in clinical practice (I, II, III, aVR, aVL, aVF, and V), may confuse annotators and create bias by presenting unfamiliar ECG leads.	**Avoid Research Derived Chest Leads:** Do not use chest leads in annotation tool or use chest leads as a secondary assessment displayed below bedside view leads (I, II, III, aVR, aVL, aVF, and the V lead).
Underestimation of Time Required for Annotation: Excessive time spent on each annotation by annotators when deliberating on true/false.	1. Conduct **Pilot Study** followed by clarification of disagreements with an ECG expert. 2. Time the pilot study to approximate seconds required per annotation.
ECG Reviewer Expertise Variance: Variation in annotation depending on level of expertise for each annotator.	Validate the level of expertise for annotators by assessing qualifications as well as in-depth analysis of annotations in a significant, diverse pilot study prior to any training to determine ECG skill level. Test annotators prior to annotation to ensure required skills are met.
Allow Annotator to Determine Order of Annotation: Pilot testing and intermittent checks of inter-rater reliability are done more efficiently if both annotators are annotating the same alarms.	**Systematized Ordering of Data:** Design an application to standardize order of alarms presented in identical order for each annotator in order to examine inter-rater reliability and time per annotation.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

6/7/19

Date